

## ITERATIVE DETECTION IN MIMO SYSTEMS

### CROSS-REFERENCE TO RELATED APPLICATIONS

Priority is claimed from provisional applications: Appl.Nos. 06/395,604, filed July 12, 2002, and 06/399,746, filed July 30, 2002. The following copending applications disclose related subject matter and have a common assignee: Appl.No. ..., filed ...(TI-34475)

### BACKGROUND OF THE INVENTION

The present invention relates to communication systems, and more particularly to multiple-input multiple-output wireless systems.

Wireless communication systems typically use band-limited channels with time-varying (unknown) distortion and may have multi-users (such as multiple cellphone users within a cell). This leads to intersymbol interference and multi-user interference, and requires interference-resistant detection for systems which are interference limited. Interference-limited systems include multi-antenna systems with multi-stream or space-time coding which have spatial interference, multi-tone systems, TDMA systems having frequency selective channels with long impulse responses leading to intersymbol interference, CDMA systems with multi-user interference arising from loss of orthogonality of spreading codes, high data rate CDMA which in addition to multi-user interference also has intersymbol interference.

Interference-resistant detectors commonly invoke one of three types of equalization to combat the interference: maximum likelihood sequence estimation, (adaptive) linear filtering, and decision-feedback equalization. However, maximum likelihood sequence estimation has problems including impractically large computation complexity for systems with multiple transmit antennas and multiple receive antennas. Linear filtering equalization, such as linear zero-forcing and linear minimum square error equalization, has low computational complexity but has relatively poor performance due to excessive noise enhancement. And decision-feedback (iterative) detectors, such as

iterative zero-forcing and iterative minimum mean square error, have moderate computational complexity but only moderate performance.

Thus there is a problem of trade-off of performance and computational complexity for multiple-input, multiple-output systems.

## SUMMARY OF THE INVENTION

The present invention provides a method and detector for multiple-input, multiple-output (MIMO) systems which iteratively detects subsets of symbols and thereby allows reduced-state maximum likelihood decisions.

This has advantages including improved performance with moderate increase in computational complexity.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1a-1b are flow and functional block diagrams.

Figures 2a-2c illustrate receivers and transmitters.

Figures 3a-3d are simulation results.

Figures 4a-4b show Turbo coding.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

### 1. Overview

Preferred embodiment detectors and detection methods for multi-input, multi-output (MIMO) systems partition the set of transmitted symbols into subsets and jointly detect the symbols within a subset and iteratively detect the subsets. This provides interference cancellation of previously detected symbols. The joint detection within a subset may be a linear soft detection followed by maximum likelihood (ML) hard decision. Figure 1a is a flow diagram, and Figure 1b illustrates functional blocks of a preferred embodiment which partitions a set of  $P$  symbol indices,  $\{1, 2, \dots, P\}$ , into  $M$  subsets,  $I_1, I_2, \dots, I_M$ , of sizes  $P_1, P_2, \dots, P_M$ , respectively. Thus,  $P = \sum_{1 \leq m \leq M} P_m$ . The hard estimates for the  $P_m$  symbols with indices in subset  $I_m$  are computed by a  $P_m$ -input,  $P_m$ -output ML detector. These hard estimates are then used to regenerate the signal components associated with the symbols whose indices lie in  $I_m$ , and the regenerated components from this iteration plus prior iterations are subtracted from the received signal (interference cancellation) for use as input to the next iteration. This use of ML detection in reduced states (subsets of indices) provides a range of performance-complexity trade-offs previously unavailable.

Preferred embodiment wireless communication systems incorporate preferred embodiment detection methods. Figure 1b illustrates functional blocks of a preferred embodiment detector. Figure 2a shows a corresponding receiver for a MIMO wireless communications system; and Figure 2b shows a  $P$ -antenna transmitter for such a system. In such systems components, base stations, and mobile users, could each include one or more application specific integrated circuits (ASICs), (programmable) digital signal processors (DSPs), and/or other programmable devices with stored programs for implementation of the preferred embodiment. The base stations and mobile users may also contain analog integrated circuits for amplification of inputs to or outputs from antennas and conversion between analog and digital; and these analog and processor circuits may be integrated on a single die. The stored programs may, for example, be in

external or onboard ROM, flash EEPROM, and/or FeRAM. The antennas may be parts of RAKE detectors with multiple fingers for each user's signals. The DSP core could be a TMS320C6xxx or TMS320C5xxx from Texas Instruments.

## 2. Iterative detectors

First consider iterative detectors generally. Figure 2a illustrates a receiver with an interference-resistant detector as could be used in a wireless communications system with  $P$  transmit antennas ( $P$  data streams) and  $Q$  receive antennas. Figure 2b illustrates a corresponding transmitter with  $P$  transmit antennas. The received signal in such a system can be written as:

$$\mathbf{r} = \mathbf{H} \mathbf{s} + \mathbf{w}$$

where  $\mathbf{r}$  is the  $Q$ -vector of samples of the received baseband signal (complex numbers) corresponding to a transmission time  $n$ :

$$\mathbf{r} = \begin{bmatrix} r_1(n) \\ r_2(n) \\ \vdots \\ r_Q(n) \end{bmatrix};$$

$\mathbf{s}$  is the  $P$ -vector of transmitted symbols (complex numbers of a symbol constellation) for time  $n$ :

$$\mathbf{s} = \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_P(n) \end{bmatrix};$$

$\mathbf{H}$  is the  $Q \times P$  channel matrix of attenuations and phase shifts; and  $\mathbf{w}$  is a  $Q$ -vector of samples of received (white) noise. That is, the  $(q,p)$ th element of  $\mathbf{H}$  is the channel (including multipath combining and equalization) from the  $p$ th transmit source to the  $q$ th receive sink, and the  $q$ th element of  $\mathbf{w}$  is the noise seen at the  $q$ th receive sink.

Note that the foregoing relation applies generally to various systems with various interference problems and in which  $n$ ,  $\mathbf{r}$ ,  $\mathbf{s}$ ,  $P$ , and  $Q$  have corresponding interpretations. For example:

(a) High data rate multi-antenna systems such as BLAST (Bell Labs layered space time) or MIMO and multi-stream space-time coding: spatial interference suppression techniques are used in detection.

(b) Broadband wireless systems employing OFDM (orthogonal frequency division multiplex) signaling and MIMO techniques for each tone or across tones.

(c) TDMA (time division multiple access) systems having frequency-selective channels with long impulse response which causes severe ISI (intersymbol interference). Use equalizers to mitigate ISI.

(d) CDMA (code division multiple access) systems having frequency-selective channels which cause MUI (multi-user interference) as a result of the loss of orthogonality between spreading codes. For high data rate CDMA systems such as HSDPA and 1xEV-DV, this problem is more severe due to the presence of ISI. Equalizers and/or interference cancellation may be used to mitigate these impairments.

(e) Combinations of foregoing.

Hence,  $P$  is essentially the number of symbols that are jointly detected as they interfere with one another, and  $Q$  is simply the number of collected samples at the receiver. Because there are  $P$  independent sources,  $Q$  must be at least as large as  $P$  to separate the  $P$  symbols. A (interference-resistant) detector in a receiver as in Figures 2a-2b analyzes received signal vector  $\mathbf{r}$  and outputs soft estimates  $\mathbf{x}$  of the transmitted symbols  $\mathbf{s}$  to a demodulator and decoder.

Presume that the different symbols which are transmitted via  $P$  different antennas are uncorrelated and also may utilize different modulation schemes. This implies the  $P \times P$  matrix of expected symbol correlations,  $\mathbf{\Lambda} = E[\mathbf{s}\mathbf{s}^H]$ , is diagonal with entries equal the expected symbol energies ( $\lambda_k = E[|s_k|^2]$ ):

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_P \end{bmatrix}$$

For linear filtering equalization detectors, such as linear zero-forcing (LZF) or linear minimum mean square error (LMMSE), the soft estimates, denoted by  $P$ -vector  $\mathbf{x}$ , derive from the received signal by linear filtering with  $P \times Q$  matrix  $\mathbf{F}$ ; namely,  $\mathbf{x} = \mathbf{F} \mathbf{r}$ .

In particular, LMMSE detection finds the matrix  $\mathbf{F}$  by minimizing the mean square error,  $E[||\mathbf{x} - \mathbf{s}||^2]$ . With perfect estimation of the channel  $\mathbf{H}$ , the minimizing matrix  $\mathbf{F}$  is given by :

$$\begin{aligned}\mathbf{F}_{LMMSE} &= [\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{\Lambda}^{-1}]^{-1} \mathbf{H}^H \\ &= \mathbf{\Lambda} \mathbf{H}^H [\mathbf{H} \mathbf{\Lambda} \mathbf{H}^H + \sigma^2 \mathbf{I}_Q]^{-1}\end{aligned}$$

where  $\sigma^2$  is the variance per symbol of the additive white noise  $\mathbf{w}$  and  $\mathbf{I}_Q$  is the  $Q \times Q$  identity matrix. Note  $\mathbf{F}$  has the form of a product of an equalization matrix with  $\mathbf{H}^H$  which is the matrix of the matched filter for the channel.

Analogously, LZF detection finds the matrix  $\mathbf{F}$  by inversion:

$$\mathbf{F}_{LZF} = [\mathbf{H}^H \mathbf{H}]^{-1} \mathbf{H}^H$$

Essentially, LZF detection nulls out the interference at the expense of excessive noise enhancement, while LMMSE detection minimizes the sum of interference and noise energy to provide a balanced trade-off between bias (due to residual interference) and variance (due to noise enhancement). But then each component of  $\mathbf{x}$  is processed separately to obtain a corresponding hard symbol estimate. That is, a hard estimate,  $\hat{s}_i$ , for the  $i$ th symbol,  $s_i$ , comes from application of a hard decision operator to the soft estimate:  $\hat{s}_i = \mathcal{D}\{x_i\}$  for  $i = 1, 2, \dots, P$ , where  $\mathcal{D}\{\}$  is a non-linear decision device, usually a single-input, single-output (SISO) maximum-likelihood (ML) detector.

A typical iterative (decision-feedback) detector (IZF or IMMSE) for blocks of  $P$  symbols has a series of  $P$  linear detectors ( $P$  iterations) with each linear detector followed by a (hard) decision device and interference subtraction (cancellation). Each of the  $P$  linear detectors (iterations) generates both a hard and a soft estimate for one of the  $P$  symbols. The hard estimate is used to regenerate the interference arising from the already-estimated symbols which is then subtracted from the received signal, and the difference used for the next

linear symbol estimation. Hence, the hard symbol estimates,  $\hat{s}_i$ , are also computed in a disjoint fashion.

On the other hand, the optimal joint maximum likelihood (ML) detector generates the hard symbol estimates as:

$$\hat{\mathbf{s}}_{ML} = \arg \min_{\mathbf{s} \in C} \|\mathbf{r} - \mathbf{H}\mathbf{s}\|^2$$

where  $C = C_1 \times C_2 \times \dots \times C_P$  is the  $P$ -dimensional constellation of  $P$ -vectors of symbols and has size  $\prod_{1 \leq j \leq P} |C_j|$  where  $|C_j|$  denotes the number of elements in constellation  $C_j$ . Thus the hard symbol estimates are jointly computed, and the computational complexity of this optimal detection is proportional to the constellation size which is exponential in  $P$ .

When an error correcting code is employed, soft symbol or soft bit (log likelihood ratio) estimates are often desired (see Section 8 for example). While soft symbol estimates can be obtained in many ways, the conditional expectation estimator is optimal. The conditional expectation estimator is:

$$\hat{\mathbf{s}} = E[\mathbf{s} | \mathbf{r}] = \frac{\sum_{\mathbf{s} \in C} \mathbf{s} \times \exp(-\|\mathbf{r} - \mathbf{H}\mathbf{s}\|^2 / \sigma^2)}{\sum_{\mathbf{s} \in C} \exp(-\|\mathbf{r} - \mathbf{H}\mathbf{s}\|^2 / \sigma^2)}$$

The performance loss of iterative detectors comes from the disjoint processing of the soft symbol estimates,  $x_i$ , to obtain the corresponding hard symbol estimates,  $\hat{s}_i$ . The preferred embodiment detectors circumvent this problem by partial ML detections with moderate computational complexity.

### 3. Iterative reduced-state maximum-likelihood detection

Figure 1a is a flow diagram and Figure 1b shows functional blocks for first preferred embodiment detection methods which detect a  $P$ -vector of transmitted symbols,  $\mathbf{s}$ , from a received  $Q$ -vector signal,  $\mathbf{r}$ , as expressed in the following pseudocode which presumes a partition of the set of symbol indices,  $\{1, 2, \dots, P\}$ , into  $M$  subsets of indices  $I_1, I_2, \dots, I_M$ , where  $I_m$  has  $P_m$  elements.

initialize  $\mathbf{r}^{(1)} = \mathbf{r}$

for  $m = 1$  to  $M$

$\mathbf{x}^{(m)} = \mathbf{F}_m \mathbf{r}^{(m)}$  //  $P_m$ -vector of soft estimates for  $I_m$  indexed symbols

$\hat{\mathbf{s}}^{(m)} = \mathbf{MLD}_m \{ \mathbf{x}^{(m)} \}$  //  $P_m$ -vector of hard decisions for  $I_m$  indexed symbols

$\mathbf{r}^{(m+1)} = \mathbf{r}^{(m)} - \mathbf{G}_m \hat{\mathbf{s}}^{(m)}$  // interference cancellation using hard estimates

end

where  $\mathbf{F}_m$  is a  $P_m \times Q$  matrix for the feedforward transformation of received signal into soft estimates,  $\mathbf{x}^{(m)}$ , of the  $I_m$  indexed symbols;  $\mathbf{MLD}_m$  is a  $(P_m, P_m)$  maximum likelihood (ML) detector;  $\mathbf{G}_m$  is a  $Q \times P_m$  matrix for the feedback transformation using the  $m$ th iteration output for interference cancellation. The following paragraphs describe choices for  $\mathbf{F}_m$ ,  $\mathbf{G}_m$ , and  $\mathbf{MLD}_m$ . And note that the preferred embodiment methods for trivial index set partitions reduce to the usual methods: for  $M = 1$  (no partitioning), the method is just an ML detection for all of the symbols, and for  $M = P$  (every index is a separate subset), the method is just regular iterative detection.

Zero-forcing and minimum mean square error criteria each define  $\mathbf{F}_m$  and  $\mathbf{G}_m$  as follows. Begin with some convenient notation: let  $J_1 = \{1, 2, \dots, P\}$  and recursively take  $J_m = J_{m-1} - I_{m-1}$  for  $2 \leq m \leq M$ , so  $J_m$  is the set of indices of symbols remaining to be detected at the start of the  $m$ th iteration. Let  $\mathbf{H}_m$  be the  $Q \times P_m$  matrix consisting of the columns of  $\mathbf{H}$  with index in  $I_m$ , so  $\mathbf{H}_m$  is the channel matrix for the  $P_m$  symbols being detected during the  $m$ th iteration. Let  $\mathbf{K}_m$  be the  $Q \times (\sum_{m \leq k \leq M} P_k)$  matrix consisting of the columns of  $\mathbf{H}$  with index in  $J_m$ , so  $\mathbf{K}_m$  is the channel matrix for the symbols remaining to be detected at the start of the  $m$ th iteration ( $\mathbf{K}_1 = \mathbf{H}$ ). Let  $\mathbf{\Lambda}_m$  be the  $P_m \times P_m$  diagonal sub-matrix of  $\mathbf{\Lambda}$  of elements with both row and column indices in  $I_m$ ; consequently, the elements of  $\mathbf{\Lambda}_m$  are the energies of the symbols to be detected during the  $m$ th iteration. Let  $\mathbf{\Pi}_m$  be the  $(\sum_{m \leq k \leq M} P_k) \times (\sum_{m \leq k \leq M} P_k)$  diagonal sub-matrix of  $\mathbf{\Lambda}$  of elements with both row and column indices in  $J_m$ , so the elements of  $\mathbf{\Pi}_m$  are the energies of the symbols remaining to be detected at the start of the  $m$ th iteration. Let  $\mathbf{1}_m$  be the  $P_m \times Q$  the matrix of the rows of the identity matrix  $\mathbf{I}_Q$  with row index in  $I_m$ ; that is,  $\mathbf{1}_{m(p,q)} = 1$  if



$q$  is the  $p$ th element of  $I_m$  and  $= 0$  otherwise, and thus  $\mathbf{1}_m$  extracts the subspace spanned by the symbols to be detected in the  $m$ th iteration.

Neglecting decision feedback error and choosing  $\mathbf{F}_m$  and  $\mathbf{G}_m$  according to the zero-forcing criterion yields for  $m = 1, 2, \dots, M-1$ :

$$\begin{aligned}\mathbf{F}_m &= \mathbf{1}_m [\mathbf{K}_m^H \mathbf{K}_m]^{-1} \mathbf{K}_m^H \\ \mathbf{G}_m &= \mathbf{H}_m\end{aligned}$$

Analogously, choosing  $\mathbf{F}_m$  and  $\mathbf{G}_m$  according to the minimum mean square error criterion yields for  $m = 1, 2, \dots, M-1$ :

$$\begin{aligned}\mathbf{F}_m &= \mathbf{1}_m [\mathbf{K}_m^H \mathbf{K}_m + \sigma^2 \mathbf{\Lambda}_m^{-1}]^{-1} \mathbf{K}_m^H \\ &= \mathbf{\Lambda}_m \mathbf{H}_m^H [\mathbf{K}_m \mathbf{\Lambda}_m \mathbf{K}_m^H + \sigma^2 \mathbf{I}_Q]^{-1} \\ \mathbf{G}_m &= \mathbf{H}_m\end{aligned}$$

In both cases during the last iteration ( $m = M$ ) all remaining symbols will be detected, so  $J_M = I_M$ ,  $\mathbf{K}_M = \mathbf{H}_M$ ,  $\mathbf{\Lambda}_M = \mathbf{\Lambda}_M$ . Further, there is no next iteration and thus no need for a cancellation matrix  $\mathbf{G}_M$ ; see Figure 1b. The following sections 4 and 5 provide an explicit example and an extension to compensate for decision feedback error.

The maximum likelihood hard decision for the  $m$ th iteration,  $\mathbf{MLD}_m$ , depends upon the signal model for  $\mathbf{r}^{(m)}$ . In particular,

$$\begin{aligned}\mathbf{r}^{(m)} &= \mathbf{r}^{(m-1)} - \mathbf{G}_{m-1} \hat{\mathbf{s}}^{(m-1)} \\ &= \mathbf{r} - \sum_{1 \leq k \leq m-1} \mathbf{G}_k \hat{\mathbf{s}}^{(k)} \\ &= \mathbf{H}\mathbf{s} + \mathbf{w} - \sum_{1 \leq k \leq m-1} \mathbf{G}_k \hat{\mathbf{s}}^{(k)}\end{aligned}$$

where  $\hat{\mathbf{s}}^{(k)}$  is the  $P_k$ -vector of hard estimates of  $\mathbf{s}^{(k)}$  which is the  $P_k$ -vector of symbols with indices in subset  $I_k$ .

Now for the case of no feedback error,  $\mathbf{G}_k = \mathbf{H}_k$  and  $\hat{\mathbf{s}}^{(k)} = \mathbf{s}^{(k)}$ , so the model becomes

$$\mathbf{r}^{(m)} = \sum_{m \leq j \leq M} \mathbf{H}_j \mathbf{s}^{(j)} + \mathbf{w}$$

Thus the soft estimates can be written as:

$$\begin{aligned}\mathbf{x}^{(m)} &= \mathbf{F}_m \mathbf{r}^{(m)} \\ &= \mathbf{F}_m \mathbf{H}_m \mathbf{s}^{(m)} + \sum_{m+1 \leq j \leq M} \mathbf{F}_m \mathbf{H}_j \mathbf{s}^{(j)} + \mathbf{F}_m \mathbf{w} \\ &= \mathbf{F}_m \mathbf{H}_m \mathbf{s}^{(m)} + \mathbf{i}_m\end{aligned}$$

where  $\mathbf{i}_m$  is the residual interference-plus-noise term corresponding to the  $m$ th iteration detection. Note that the desired signal term,  $\mathbf{F}_m \mathbf{H}_m \mathbf{s}^{(m)}$ , is uncorrelated to the residual interference-plus-noise and that

$$\mathbb{E}[\mathbf{i}_m \mathbf{i}_m^H] = \mathbf{F}_m \mathbf{K}_{m+1} \mathbf{\Lambda}_m (\mathbf{F}_m \mathbf{K}_{m+1})^H + \sigma^2 \mathbf{F}_m \mathbf{F}_m^H$$

For notational convenience define  $\mathbf{\Psi}_m = \mathbb{E}[\mathbf{i}_m \mathbf{i}_m^H]$  which is a  $P_m \times P_m$  matrix.

Then assuming the residual interference is approximately Gaussian with covariance matrix  $\mathbf{\Psi}_m$ , so log-likelihood metric for  $\mathbf{MLD}_m$  is just the Euclidean metric scaled by the covariance matrix:

$$\hat{\mathbf{s}}^{(m)} = \arg \min_{\mathbf{c}^{(m)}} \|\mathbf{\Psi}_m^{-1/2} (\mathbf{x}^{(m)} - \mathbf{F}_m \mathbf{H}_m \mathbf{c}^{(m)})\|^2$$

where  $\mathbf{c}^{(m)}$  is a  $P_m$ -vector of symbols in the product of the  $P_m$  constellations of symbols with indices in  $I_m$ . Thus the hard decision is found by computing all of the metrics and taking the symbol vector which generates the minimum metric; and if all of the constellations have the same size  $|C|$  (such as all are  $|C|$ -QAM), then the number of metrics to compute is  $|C|^{P_m}$ .

Note that for  $\mathbf{F}_m$  and  $\mathbf{G}_m$  defined using the zero-forcing criterion there is no residual interference because  $\mathbf{F}_m \mathbf{K}_{m+1} = \mathbf{0}$ . In this case  $\mathbf{i}_m$  is exactly Gaussian. For the MMSE criterion, however, some residual interference exists; hence,  $\mathbf{i}_m$  is a mixture of Gaussian (from the noise) and non-Gaussian (residual interference) random processes. For sufficiently dense symbol constellations and/or a sufficiently large number of antennas, the interference can be well approximated as Gaussian. Otherwise, a variation of the MLD decision for MMSE-based detection is the use of the exact probability density function (pdf) of  $\mathbf{i}_m$  in the maximum likelihood detection. Because  $\mathbf{i}_m$  is the sum of  $\mathbf{F}_m \mathbf{w}$  and  $\sum_{m+1 \leq j \leq M} \mathbf{F}_m \mathbf{H}_j \mathbf{s}^{(j)}$ , its pdf is the convolution of their pdfs.

#### 4. 4x4 example

As an example consider the 4x4 MIMO system with the preferred embodiment partitioning the symbols into two subsets of two symbols each; that is,  $P = Q = 4$  with  $P_1 = P_2 = 2$  ( $M = 2$ ). Explicitly, take  $I_1 = \{1,2\}$  and  $I_2 = \{3,4\}$ ; then with  $\mathbf{h}_k$  denoting the  $k$ th column of 4x4 channel matrix  $\mathbf{H}$ , the zero-forcing criterion yields:

$$\mathbf{F}_1 = \mathbf{1}_{\{1,2\}} [\mathbf{H}^H \mathbf{H}]^{-1} \mathbf{H}^H \quad (2 \times 4 \text{ matrix})$$

$$\mathbf{G}_1 = [\mathbf{h}_1 \ \mathbf{h}_2] \quad (4 \times 2 \text{ matrix})$$

$$\mathbf{F}_2 = [\mathbf{h}_3 \ \mathbf{h}_4]^H \quad (2 \times 4 \text{ matrix})$$

Analogously, the MMSE criterion gives:

$$\mathbf{F}_1 = \mathbf{1}_{\{1,2\}} [\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I}_4]^{-1} \mathbf{H}^H \quad (2 \times 4 \text{ matrix})$$

$$= \begin{bmatrix} \lambda_1 \mathbf{h}_1^H \\ \lambda_2 \mathbf{h}_2^H \end{bmatrix} [\lambda_1 \mathbf{h}_1 \mathbf{h}_1^H + \lambda_2 \mathbf{h}_2 \mathbf{h}_2^H + \sigma^2 \mathbf{I}_4]^{-1}$$

$$\mathbf{G}_1 = [\mathbf{h}_1 \ \mathbf{h}_2] \quad (4 \times 2 \text{ matrix})$$

$$\mathbf{F}_2 = [\mathbf{h}_3 \ \mathbf{h}_4]^H \quad (2 \times 4 \text{ matrix})$$

The simulations of section 9 compare performance of various modifications of this 4x4 MIMO system preferred embodiment plus regular iterative detection.

## 5. Feedback error compensation

In general, the error-free decision feedback assumption is sufficiently accurate for high signal-to-noise ratio (SNR) and good channel conditions. Otherwise, error propagation may occur. When the MMSE criterion is used, the effect of error propagation can be taken into account by using the second order model which presumes the correlation of the transmitted symbols and their hard estimates is expressed by a diagonal  $P \times P$  energy-correlation matrix:

$$\mathbf{E}[\mathbf{s} \hat{\mathbf{s}}^H] = \begin{bmatrix} \lambda_1 \rho_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_P \rho_P \end{bmatrix}$$

where the symbol-hard estimate correlations  $\rho_j$  are generally real-valued and in the range  $-1 \leq \rho_j \leq +1$ . Let  $\mathbf{R}$  denote the  $P \times P$  diagonal matrix of these correlations. Then define the  $P_m \times P_m$  diagonal correlation submatrix  $\mathbf{R}_m$  with diagonal elements  $\rho_k$  for  $k$  in  $I_m$  and the  $(\sum_{m \leq k \leq M} P_k) \times (\sum_{m \leq k \leq M} P_k)$  diagonal correlation submatrix  $\mathbf{R}_m$  with diagonal elements  $\rho_k$  for  $k$  in  $J_m$ . Thus,  $\mathbf{R}_m$  is useful for interference cancellation at the end of the  $m$ th iteration when the  $\hat{\mathbf{s}}^{(m)}$  are available, so the  $\rho_k$  for  $k$  in  $I_m$  are generally non-zero in  $\mathbf{R}_m$ . Further,  $\mathbf{R}_m$  is useful

for interference suppression at the start of the  $m$ th iteration when the  $\hat{\mathbf{s}}^{(m)}$  are not yet available and so the  $\rho_k$  for  $k$  in  $I_m$  are set to zero in forming  $\mathbf{R}_m$ . All the  $\rho_k$  for  $k$  in  $J_m$  are not yet available so this is all 0s. For multistage detection, some  $\rho_k$  will be 0 for the undetected symbols in the first stage; during the second stage estimates from the previous stage for all the symbols are available and  $\rho_k$  is non-zero for every  $k$ .

Taking into account the effect of error propagation in the MMSE criterion approach leads to modified  $\mathbf{F}_m$  and  $\mathbf{G}_m$ ; for  $m = 1, 2, \dots, M-1$ :

$$\begin{aligned}\mathbf{F}_m &= \mathbf{1}_m [\mathbf{K}_m^H \mathbf{K}_m (\mathbf{I} - \mathbf{R}_m^2) + \sigma^2 \mathbf{\Lambda}_m^{-1}]^{-1} \mathbf{K}_m^H \\ &= \mathbf{\Lambda}_m \mathbf{H}_m^H [\mathbf{K}_m (\mathbf{I} - \mathbf{R}_m^2) \mathbf{\Lambda}_m \mathbf{K}_m^H + \sigma^2 \mathbf{I}_Q]^{-1} \\ \mathbf{G}_m &= \mathbf{H}_m \mathbf{R}_m\end{aligned}$$

Note that these modified matrices reduce to the prior expressions when  $\rho_k = 1$  for previously-detected symbols. In particular,  $\rho_k = 1$  for  $k$  in  $J_m - I_m$  reduces  $\mathbf{F}_m$  to the prior expression and  $\rho_k = 1$  for  $k$  in  $I_m$  reduces  $\mathbf{G}_m$  to the prior expression.

This feedback error compensation by use of correlations to limit cancellation allows for a multistage detection with each stage a set of  $M$  iterations as in the foregoing and with the hard estimates and correlations of one stage used as initial conditions for the next stage. In this manner, performance can be improved in stages by controlling the cancellation according to current correlation. Of course, the correlations must be computed, either by ensemble averaging (i.e., use a pilot channel or a set of training symbols that undergo the same MIMO channel as the information symbols) or by use of an analytical model (e.g., specific symbol constellation and Gaussian approximations) which can relate correlations and SINRs (which are used to define the  $\mathbf{F}$  and  $\mathbf{G}$  matrices) plus provide convergence to limiting values through successive stages.

In particular, if superscript/subscript  $[n]$  denotes  $n$ th stage variables, then the first stage would be as described in the preceding paragraphs: the  $m$ th iteration would include:

$$\begin{aligned}\mathbf{r}^{[1](m)} &= \mathbf{r}^{[1](m-1)} - \mathbf{G}_{[1]m-1} \hat{\mathbf{s}}^{[1](m-1)} \\ &= \mathbf{r} - \sum_{1 \leq k \leq m-1} \mathbf{G}_{[1]k} \hat{\mathbf{s}}^{[1](k)}\end{aligned}$$

$$\begin{aligned}
\mathbf{F}_{[1]m} &= \mathbf{\Lambda}_m \mathbf{H}_m^H [\mathbf{K}_m (\mathbf{I} - \mathbf{\mathcal{R}}_{[1]m}^2) \mathbf{\Pi}_m \mathbf{K}_m^H + \sigma^2 \mathbf{I}_Q]^{-1} \\
\mathbf{x}^{[1](m)} &= \mathbf{F}_{[1]m} \mathbf{r}^{[1](m)} \quad \dots \\
\hat{\mathbf{s}}^{[1](m)} &= \mathbf{MLD}_m \{ \mathbf{x}^{[1](m)} \} \\
\mathbf{G}_{[1]m} &= \mathbf{H}_m \mathbf{R}_{[1]m}
\end{aligned}$$

Then the second stage would have first stage hard estimates of interfering subsets of symbols to include in the cancellations, so

$$\begin{aligned}
\mathbf{r}^{[2](m)} &= \mathbf{r}^{[2](m-1)} - \mathbf{G}_{[2]m-1} \hat{\mathbf{s}}^{[2](m-1)} \\
&= \mathbf{r} - \sum_{1 \leq k \leq m-1} \mathbf{G}_{[2]k} \hat{\mathbf{s}}^{[2](k)} - \sum_{m+1 \leq j \leq M} \mathbf{G}_{[1]j} \hat{\mathbf{s}}^{[1](j)}
\end{aligned}$$

Note that the hard estimates from the first stage for the subset of symbols being detected,  $\hat{\mathbf{s}}^{[1](m)}$ , are not part of the cancellation. And the correlation matrices  $\mathbf{\mathcal{R}}_{[2]m}$  can use first stage correlations  $\rho_{[1]k}$  for  $k$  in  $J_m$ , whereas  $\mathbf{R}_{[2]m}$  is used after the  $m$ th iteration and will have correlations  $\rho_{[2]k}$  for  $k$  in  $J_m$ . And so forth for subsequent stages.

## 6. Ordered detection

Ordered detection based on the symbol post-detection signal-to-interference-plus-noise ratio (SINR) is often used to further improve performance. error. Selecting the detection ordering in the foregoing preferred embodiments amounts to selecting the index sets  $I_1, I_2, \dots, I_M$  that form a partition of  $\{1, 2, \dots, P\}$ . One may select a pre-determined detection ordering that does not depend on the channel realization  $\mathbf{H}$ . Analogous to the conventional iterative detection, further improvement can be obtained by selecting the ordering as a function of  $\mathbf{H}$  based on a selected performance measure as in the following examples.

### (1) Pre-ML SINR measures

At the start of the  $m$ th iteration, compare the  $(\sum_{m \leq k \leq M} P_k)$  SINR values associated with the index set  $J_m$ , and select index set  $I_m$  to label the  $P_m$  largest SINR values. Note that closed form expressions for the SINR as a function of the channel realization can be obtained for different detection criteria. In particular, define the SINR associated with symbol  $s_p$  at the  $m$ th iteration as  $\gamma_p^{(m)}$ . For the zero-forcing criterion this yields for  $p \in I_m$ :

$$\gamma_p^{(m)} = \lambda_p / \sigma^2 [(\mathbf{K}_m^H \mathbf{K}_m)^{-1}]_{p,p}$$

Whereas the MMSE criterion together with the feedback error compensation yields for  $p \in I_m$ :

$$\begin{aligned}\gamma_p^{(m)} &= \lambda_p \mathbf{h}_p^H [\sum_{n \in J_{m+1}} \lambda_n (1 - \rho_n^2) \mathbf{h}_n \mathbf{h}_n^H + \sum_{k \in I_{m-p}} \lambda_k \mathbf{h}_k \mathbf{h}_k^H + \sigma^2 \mathbf{I}_Q]^{-1} \mathbf{h}_p \\ &= \lambda_p / [\sigma^2 \mathbf{\Lambda}_m^{-1} (\mathbf{K}_m^H \mathbf{K}_m) (\mathbf{I} - \mathbf{\Lambda}_m^2) + \sigma^2 \mathbf{\Lambda}_m^{-1}]_{p,p}^{-1} - 1\end{aligned}$$

where  $\mathbf{h}_k$  is the  $k$ th column of  $\mathbf{H}$ , and the case of error-free feedback is a special case.

## (2) Post-ML SINR measures

Intuitively, better detection ordering can be obtained when a measure of SINR after the ML detection. Several possible post-ML detection measures can be derived from the symbol error rate (SER) expression of ML detection. Starting from the ML metric above, scale by  $\Psi_m^{-1/2}$  to define:

$$\begin{aligned}\mathbf{y}^{(m)} &\stackrel{\text{def}}{=} \Psi_m^{-1/2} \mathbf{x}^{(m)} \\ \mathbf{\Theta}_m &\stackrel{\text{def}}{=} \Psi_m^{-1/2} \mathbf{F}_m \mathbf{K}_m \\ \mathbf{j}_m &\stackrel{\text{def}}{=} \Psi_m^{-1/2} \mathbf{i}_m\end{aligned}$$

Then

$$\begin{aligned}\mathbf{y}^{(m)} &= \Psi_m^{-1/2} \mathbf{x}^{(m)} \\ &= \Psi_m^{-1/2} (\mathbf{F}_m \mathbf{K}_m \mathbf{s}^{(m)} + \mathbf{i}_m) \\ &= \Psi_m^{-1/2} \mathbf{\Theta}_m \mathbf{s}^{(m)} + \mathbf{j}_m\end{aligned}$$

Here the scaled residual interference-plus-noise term,  $\mathbf{j}_m$ , is white with  $E[\mathbf{j}_m \mathbf{j}_m^H] = \mathbf{I}$ . Let  $\mathcal{S}_m$  denote the  $P_m$ -dimensional product of constellations of the symbols with indices in  $I_m$ ; that is,  $\mathcal{S}_m = \prod_{k \in I_m} C_k$  where  $C_k$  is the symbol constellation for  $s_k$  and the product is over  $k$  in  $I_m$ . Also let  $N_m$  denote the size of  $\mathcal{S}_m$ , so  $N_m = \prod_{k \in I_m} |C_k|$ . An upper bound of symbol error rate (SER) corresponding to  $\mathbf{MLD}_m$  can be derived as follows:

$$\begin{aligned}SER_m &= \Pr(\hat{\mathbf{s}}^{(m)} \neq \mathbf{s}^{(m)}) \\ &= (1/N_m) \sum_{\alpha \in \mathcal{S}_m} \Pr(\hat{\mathbf{s}}^{(m)} \neq \mathbf{s}^{(m)} | \mathbf{s}^{(m)} = \alpha) \\ &= (1/N_m) \sum_{\alpha \in \mathcal{S}_m} \sum_{\hat{\alpha}, \hat{\alpha} \neq \alpha} \Pr(\hat{\mathbf{s}}^{(m)} = \hat{\alpha} | \mathbf{s}^{(m)} = \alpha)\end{aligned}$$

$$\begin{aligned}
&= (1/ N_m) \sum_{\alpha \in \mathcal{S}_m} \sum_{\dot{\alpha}, \dot{\alpha} \neq \alpha} \Pr(\bigcap_{\beta \neq \dot{\alpha}} \{ \|\mathbf{y}^{(m)} - \Theta_m \dot{\alpha}\|^2 \leq \|\mathbf{y}^{(m)} - \Theta_m \beta\|^2 \}) \\
&\leq (1/ N_m) \sum_{\alpha \in \mathcal{S}_m} \sum_{\dot{\alpha}, \dot{\alpha} \neq \alpha} \Pr(\|\mathbf{y}^{(m)} - \Theta_m \dot{\alpha}\|^2 \leq \|\mathbf{y}^{(m)} - \Theta_m \alpha\|^2) \\
&= (1/ N_m) \sum_{\alpha \in \mathcal{S}_m} \sum_{\dot{\alpha}, \dot{\alpha} \neq \alpha} \mathbb{Q}(\sqrt{(\|\Theta_m(\alpha - \dot{\alpha})\|^2 / 2)})
\end{aligned}$$

where the  $\mathbb{Q}$  function is the area under the Gaussian tail. This upper bound on  $SER_m$  can be used to select the detection order: take  $I_m$  to minimize the upper bound. Thus this measure is post-ML based on the SER upper bound.

A looser upper bound can be derived from the fact that for  $\alpha, \dot{\alpha} \in \mathcal{S}_m$ :

$$\|\Theta_m(\alpha - \dot{\alpha})\|^2 \geq \min_{1 \leq i \leq P_m} (\|[\Theta_m]_{:,i}\|^2 d_{min,i}^2)$$

where  $d_{min,i}^2$  is the minimum distance among symbols in the symbol constellation associated with the  $i$ th column of  $\Theta_m$ ,  $[\Theta_m]_{:,i}$ . Thus,

$$\begin{aligned}
SER_m &\leq (1/ N_m) \sum_{\alpha \in \mathcal{S}_m} \sum_{\dot{\alpha}, \dot{\alpha} \neq \alpha} \mathbb{Q}(\sqrt{(\|\Theta_m(\alpha - \dot{\alpha})\|^2 / 2)}) \\
&\leq (N_m - 1) \mathbb{Q}(\sqrt{(\min_{1 \leq i \leq P_m} (\|[\Theta_m]_{:,i}\|^2 d_{min,i}^2) )})
\end{aligned}$$

When the same constellation (modulation scheme) is used for all of the transmit antennas, the minimum distance is common and the upper bound simplifies to:

$$SER_m \leq (N_m - 1) \mathbb{Q}(\sqrt{(d_{min}^2 \min_{1 \leq i \leq P_m} (\|[\Theta_m]_{:,i}\|^2) )})$$

This upper bound yields the following post-ML measure for selecting the detection order: at the  $m$ th iteration, select  $I_m$  to maximize  $\min_{1 \leq i \leq P_m} (\|[\Theta_m]_{:,i}\|^2)$ . This measure is thus post-ML based on the channel norm after MMSE transformation.

## 7. Decision alternatives

The foregoing preferred embodiments hard decision methods, **MLD<sub>m</sub>**, convert  $\mathbf{x}_k^{(m)}$  into  $\hat{\mathbf{s}}_k^{(m)}$  use a maximum likelihood approach which applies a hard-limiter (hard-clipping) non-linearity. Alternative non-linearities, generically denoted as  $\hat{\mathbf{s}}^{(m)} = \mathcal{D}\{\mathbf{x}^{(m)}\}$ , such as hyperbolic tangent, may result in better performance. In this case the decision statistics that are close to zero are

effectively scaled instead of clipped; the scaling depends upon the variance of the residual interference plus noise. For example, for a single BPSK symbol:

$$\text{Hyperbolic tangent:} \quad \mathcal{D}_{\tanh}(z) = \tanh(\text{Re}\{z\}/\xi)$$

$$\begin{aligned} \text{Soft linear clipped:} \quad \mathcal{D}_{\text{SL}}(z) &= +1 && \text{if } \text{Re}\{z\} > \xi \\ &= -1 && \text{if } \text{Re}\{z\} < -\xi \\ &= \text{Re}\{z\} && \text{if } |\text{Re}\{z\}| \leq \xi \end{aligned}$$

where  $\xi$  is a non-negative constant. For higher order modulation, the decision is done in the bit level following the bit log likelihood ratio computation (as described in Section 8). Alternatively, for QPSK, the real and imaginary parts of  $\mathcal{D}(z)$  can be taken as  $\tanh(\text{Re}\{z\}/\xi)$  and  $\tanh(\text{Im}\{z\}/\xi)$ , respectively; or the corresponding soft linear with  $\text{Re}\{z\}$  and  $\text{Im}\{z\}$ , respectively. Further, for higher order (larger constellations) the soft linear extends by selecting the constant to make  $\mathcal{D}(z) = z$  for  $z$  in the constellation. And for two or higher-dimensional constellations, extend by components. For example, with a subset of symbols of size  $P_m = 2$ , the ML decision on the soft estimate 2-vector  $\mathbf{x}^{(m)}$  gives the 2-vector of symbols in the product constellation closest to  $\mathbf{x}^{(m)}$ ; whereas, the soft linear would give the 2-vector  $[\mathcal{D}_{\text{SL}}(x_1^{(m)}), \mathcal{D}_{\text{SL}}(x_2^{(m)})]$ .

While any of those non-linear estimators can be used, the optimal soft estimator is the conditional expectation estimator (analogous to Section 2), which is

$$\hat{\mathbf{s}}^{(m)} = E[\mathbf{s}^{(m)} | \mathbf{x}^{(m)}] = \frac{\sum_{\mathbf{c}} \mathbf{c} \times \exp\left(-\|\Psi_m^{-1/2}(\mathbf{x}^{(m)} - \mathbf{F}_m \mathbf{H}_m \mathbf{c})\|^2\right)}{\sum_{\mathbf{c}} \exp\left(-\|\Psi_m^{-1/2}(\mathbf{x}^{(m)} - \mathbf{F}_m \mathbf{H}_m \mathbf{c})\|^2\right)}$$

where the summation is over the set formed by the product of  $P_m$  constellations of symbols with indices in  $I_m$  (see discussion at the end of Section 3).

For systems with large constellation and/or a number of antennas, the computational burden may be prohibitive as it increases exponentially with the number of transmit antennas. In this case, sphere decoding type algorithm can



be used to reduce computational complexity at the expense of some performance (while retaining most of the gain of maximum likelihood detector).

## 8. Decoding

The demodulator converts the soft symbol estimates  $x^{(1)}, x^{(2)}, \dots, x^{(M)}$  output by an M-iteration detector into conditional probabilities; and the conditional probabilities translate into bit-level log likelihood ratios (LLRs) for (sequence) decoding. In more detail, the LLRs in terms of the bits  $u_{pk}$  which define the constellation symbols  $s_p$  (e.g., two bits for a QPSK symbol, four bits for a 16QAM symbol, etc.) are defined (dropping the superscripts on the  $x_p$ ) as

$$\begin{aligned} \text{LLR}(u_{pk}) &= \log\{ P[u_{pk} = 1 \mid x_{pk}] / P[u_{pk} = 0 \mid x_{pk}] \} \\ &= \log\{ P[u_{pk} = 1 \mid x_{pk}] \} - \log\{ P[u_{pk} = 0 \mid x_{pk}] \} \end{aligned}$$

The LLRs can be computed using a channel model. For example,

$$\begin{aligned} \text{LLR}(u_{pk}) &= \log\{ P[u_{pk} = 1 \mid x_p] / P[u_{pk} = 0 \mid x_p] \} \\ &= \log\{ p(x_p \mid u_{pk} = 1) / p(x_p \mid u_{pk} = 0) \} + \log\{ P[u_{pk} = 1] / P[u_{pk} = 0] \} \end{aligned}$$

where the first log term includes the probability distribution of the demodulated symbol  $x_p$  which can be computed using the channel model. The second log term is the log of the ratio of a priori probabilities of the bit values and typically equals 0. So for an AWGN channel where the residual interference (interference which is not cancelled) is also a zero-mean, normally-distributed, independent random variable, the channel model gives:

$$p(x_p \mid s_p = c) \sim \exp( - |x_p - c|^2 / \gamma_p )$$

where  $c$  is a symbol in the symbol constellation and  $\gamma_p$  is a normalization typically derived from the channel characteristics and the detector type. Indeed, the following section describes known normalizations and preferred embodiment normalizations. Of course,  $\gamma_p$  is just twice the variance of the estimation error random variable.

Then compute LLRs by using an approximation which allows direct application of the channel model. Take  $p(x_p \mid u_{pk} = 1) = p(x_p \mid s_p = c_{pk=1})$  where  $c_{pk=1}$  is the symbol in the sub-constellation of symbols with  $k$ th bit equal 1 and which is the closest to  $x_p$ ; that is,  $c_{pk=1}$  minimizes  $|x_p - c_{k=1}|^2$  for  $c_{k=1}$  a symbol in

the sub-constellation with  $k$ th bit equal to 1. Analogously for  $p(x_p | u_{pk} = 0)$  using the sub-constellation of symbols with  $k$ th bit equal 0. Then with equal a priori probabilities of the bit values and the notation subscript  $k=1$  and  $k=0$  indicating symbols with  $k$ th bit 1 and 0, respectively, the approximation yields

$$\begin{aligned} \text{LLR}(u_{pk}) &= \log\{ p(x_p | u_{pk} = 1) / p(x_p | u_{pk} = 0) \} \\ &\cong 1/\gamma_p \cdot \{ \min_{k=0} |x_p - c_{k=0}|^2 - \min_{k=1} |x_p - c_{k=1}|^2 \} \end{aligned}$$

Thus the LLR computation just searches over the two symbol sub-constellations for the minima.

The LLRs are used in decoders for error correcting codes such as Turbo codes (e.g., iterative interleaved MAP decoders with BCJR or SOVA algorithm using LLRs for each MAP) and convolutional codes (e.g. Viterbi decoders). Such decoders require soft bit statistics (in terms of LLR) from the detector to achieve their maximum performance (hard bit statistics with Hamming instead of Euclidean metrics can also be used but result in approximately 3dB loss). Alternatively, direct symbol decoding with LLRs as the conditional probability minus the a priori probability could be used. Figures 4a-4b illustrate the 3GPP Turbo encoder and an iterative decoder which includes two MAP blocks, an interleaver, a de-interleaver, and feedback for iterations.

## 9. Simulations

Simulations for 4x4 MIMO QPSK and 16QAM systems with  $P = Q = 4$  in spatially uncorrelated (IID) and correlated MIMO channels were performed to compare a preferred embodiment with other detection methods. The correlated spatial channel profile has the average and peak cross correlation of 0.16 and 0.40, respectively. Figures 3a-3d compare simulations with the following six detection methods including four preferred embodiment iterative reduced state ML variations:

- (1) IMMSE UNORD: regular iterative MMSE without detection ordering.
- (2) IMMSE ORD: regular iterative MMSE with detection ordering based on post-detection SINR.
- (3) IRSML UNORD: MMSE-based iterative reduced state ML for  $M = 2$ ,  $P_1$

=  $P_2 = 2$  and without ordering.

(4) IRSML ORD Pre-ML SINR: MMSE-based iterative reduced state ML for  $M = 2$ ,  $P_1 = P_2 = 2$  and with ordering based on SINR after MMSE transformation.

(5) IRSML ORD Post-ML Norm: MMSE-based iterative reduced state ML for  $M = 2$ ,  $P_1 = P_2 = 2$  and with ordering based on channel norm after MMSE transformation.

(6) IRSML ORD Post-ML SER: MMSE-based iterative reduced state ML for  $M = 2$ ,  $P_1 = P_2 = 2$  and with ordering based on SER bound after ML detection.

The raw bit error rate (BER) versus signal-to-noise ratio ( $E_b/N_0$ ) results for QPSK in IID and correlated channels appear in Figures 3a-3b, respectively, and for 16QAM in IID and correlated channels appear in Figures 3c-3d, respectively. Observe that the gain of IRSMLover regular IMMSE is more dramatic in spatially correlated channels and when no detection ordering is used. Notice also that IRSML with SER-based detection ordering results in a gain of 10-dB for 16QAM in correlated channels at  $BER = 10^{-3}$  compared to the regular IMMSE with detection ordering.

The computational complexity of preferred embodiment iterative reduced state ML methods can be compared to the full ML and the regular MMSE methods. In particular, consider two aspects: matrix inversion computation and hard decision estimation. For comparison, assume that the preferred embodiments use equal size subsets ( $P_m = P/M$ ) and all methods use the same constellation (size  $C$ ) for all symbols. Then the following table indicates the complexity of the methods.

Detector	Matrix inversion	Hard decision
Regular iterative	$\sim P \times P^3$	$\sim P \times C$
Maximum likelihood	none	$\sim C^P$
IRSML	$\sim M \times P^3$	$\sim M \times C^{P/M}$

Thus for the preferred embodiments with two subsets ( $M = P/2$  and  $P_m = 2$ ),

the computational complexity of the preferred embodiments is comparable to the regular iterative detector for small to moderate constellation size. For large constellation size (such as 64-QAM or 256-QAM), the complexity increase for the preferred embodiments over the regular iterative detector becomes significant. However, its performance gain is expected to become more dramatic as indicated by the simulations.